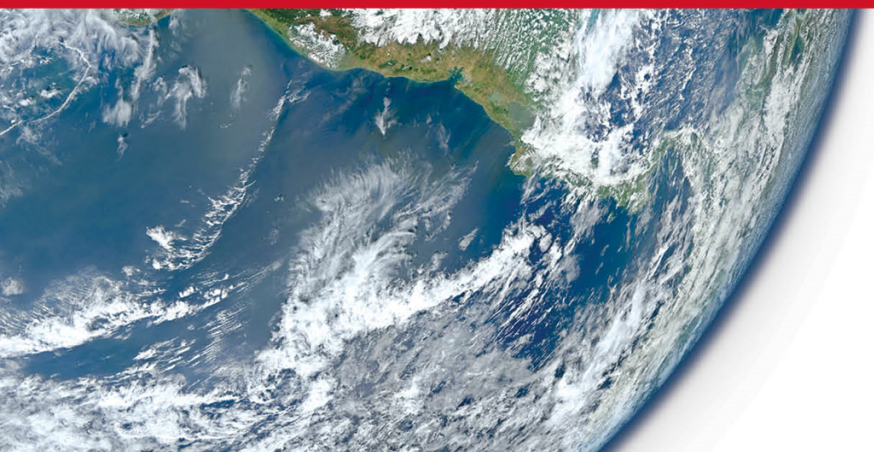
Updates on Cloud Migration and Modernization for JPSS Data Processing Node (DPN)



JPSS-CGS

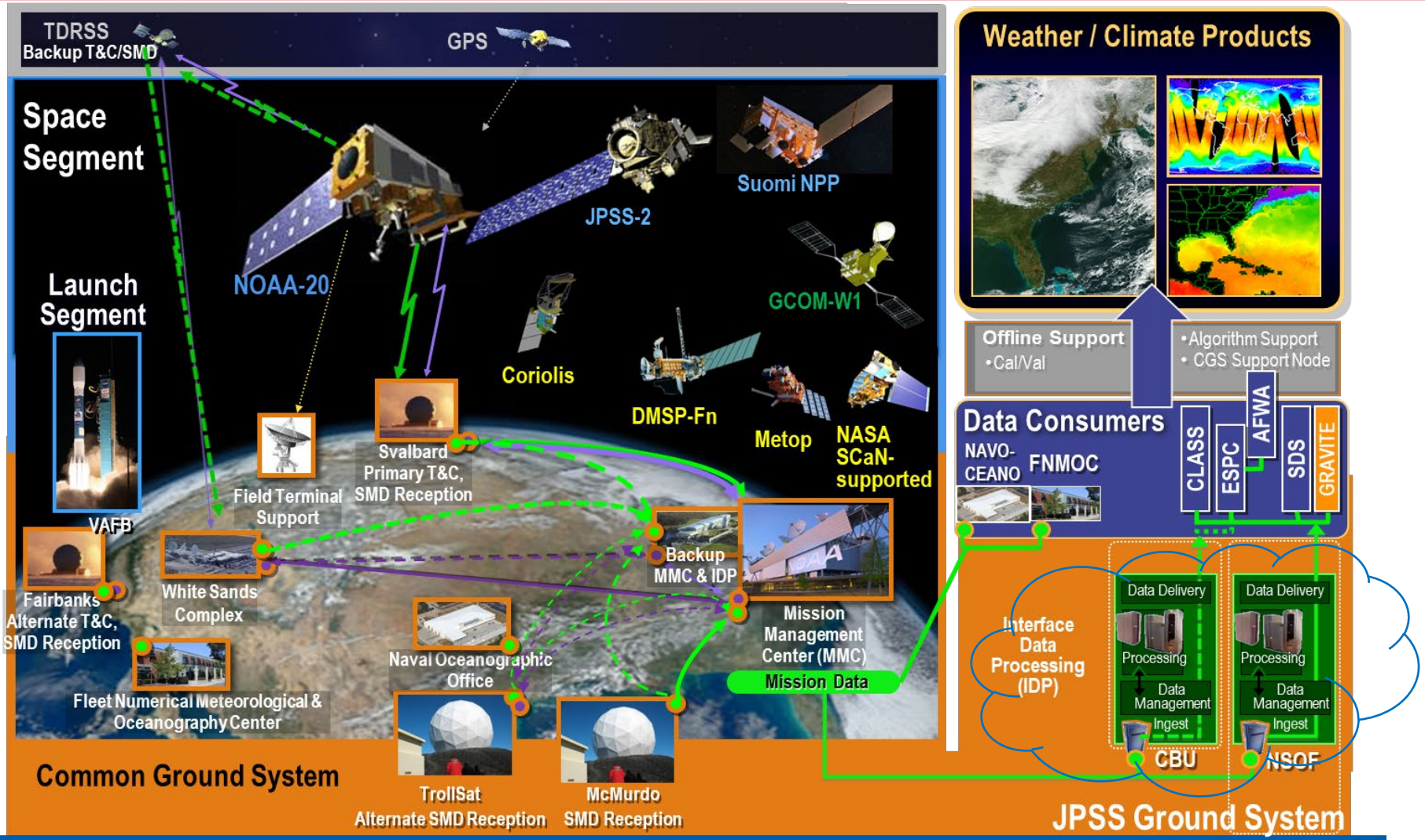
Raytheon Intelligence & Space

Jeremy Bargen

Scott Kern

Josh Olson

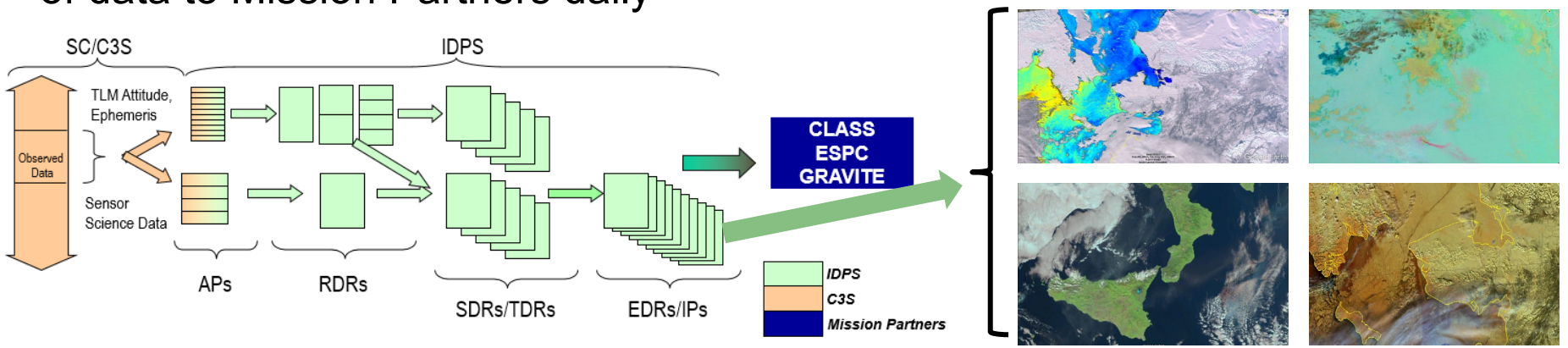
JPSS CGS Data Production – What is it?



DPN is the Joint Polar Satellite System segment which provides ground data processing to create S-NPP and JPSS data products from raw sensor data.

JPSS CGS DPN – Details

- Data Processing Node (DPN)
  - Ingests Mission Data packets (APs)
  - Produces Data Products: RDRs, SDRs, TDRs, EDRs, IPs
  - Delivers to Mission Partners: CLASS, ESPC, and GRAVITE
- Key Architectural Features
  - Configurable data driven algorithm processing chains
  - Data is processed for the S-NPP, NOAA-20 and GCOM-W1 missions
  - Migrated from on-premises operations at NOAA Satellite Operations Facility (NSOF) and Consolidated Backup (CBU) to AWS GovCloud
  - Load-balanced processing for fault management
  - NIST 800-53 v4 security implementation
  - Focus on low latency and high availability of data products
- DPN operations receives ~400 GB of data from 3 spacecraft and delivers >7 TB of data to Mission Partners daily



DPN processes observed photons into deliverable products
Observation Time to MP Delivery: <100 min

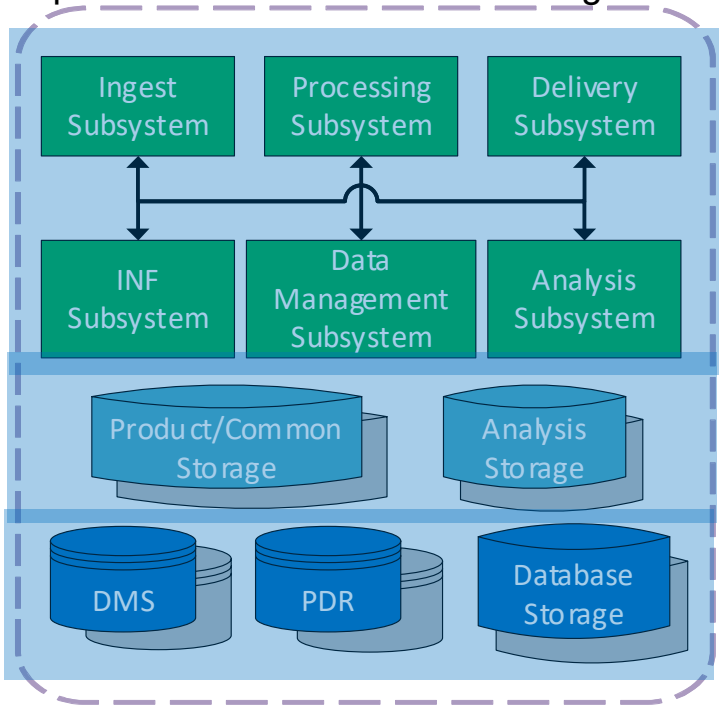
JPSS CGS DPN Cloud Migration – History and Milestones

- 2017: RTX begins evaluating cloud migration
- 2018 Part 1: Proof-of-Concept deployment to AWS
  - Initial estimate was ~1 month to get DPN running after environment configuration
  - **Completed in ~4 days!**
- 2018 Part 2: RTX DevCloud Prototype/Demos
  - Execute Trade Studies and evaluate end-to-end system performance with security tools in place
  - Demonstrated cross-AZ failover using AWS RDS DBaaS
- 2018 Part 3: MS Azure
  - Risk reduction to ensure DPN deployment to MS Azure
- 2019:
  - NOAA determined that JPSS CGS DPN will be one of the first major NOAA programs to migrate to AWS GovCloud
    - Includes dedicated environments for:
      - Operations
      - Integration and Test
      - Factory/Development
      - Algorithm Development and Assessment
- Design Review completed January 2020
- **Transition to Operations February 2021 (Phase 1 complete)**

Initial Implementation – Phase 1

- Initial migration from on-premises operational baseline to Cloud with minimal baseline changes
  - Decommission legacy hardware as quickly as possible
  - Minimize changes to baseline not explicitly necessary to operate in the cloud
    - Migrate primary DB from Oracle to PostgreSQL to reduce licensing costs
- HOT backup of primary Operations DP
  - Security Patching requires transition to backup IDP
  - 3rd IDP necessary to accommodate monthly patches and baseline upgrades while maintaining resiliency to failures
- Primary change is new "DP-Common" Environment:
  - Route data to multiple DPN systems from a single on-prem data source
  - Management of security functions
- Leveraging DevOps Tools/Processes:
  - Environments 100% managed using Infrastructure-as-Code (Packer, Terraform, Chef)
  - Faster/Frequent algorithm releases to PRO subsystem decreases Research-to-OPS (R2O) cycle
- ~60 EC2 VMs and 500 TB storage per DP

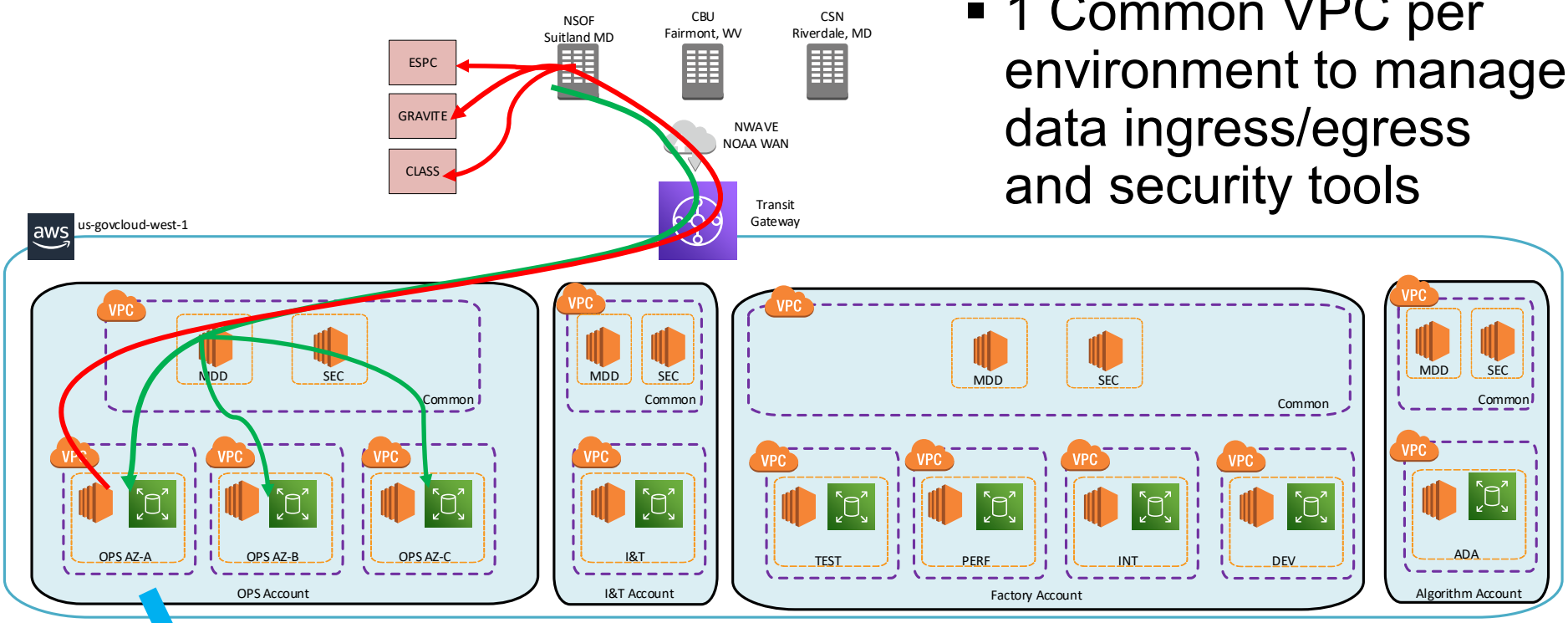
Operational Data Processor Configuration



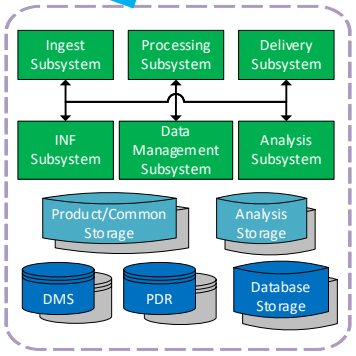
Database Layer (EC2 and EBS)
- Oracle Dataguard installed to EC2
  - Backup DB instance
- EBS storage attached to EC2

- DMS: Data Management
- PDR: Performance Data Repo

Initial Implementation – Phase 1



- 1 Common VPC per environment to manage data ingress/egress and security tools



- 9 separate data processing systems operate across 4 environments at any given time
- Future architectural simplifications and cost savings will be multiplicative

Optimization – Phase 2

- Updates the DPN cloud design to take better advantage of cloud capabilities
- Provides opportunities to significantly reduce cloud resource footprint
- Implements a better foundation for science/forecast product driven changes during Modernization Phase

Optimization	Description

Transition to Highly Available (HA) DPN	• Deploy single HA DP spanning 2 Availability Zones
	  • Subsystems deployed across AZs in auto-scaling groups
	  • "Live" security patching on dynamic instances to eliminate OPS/Non-OPS transitions for monthly security patching

Dynamic Allocation of Processing Capacity	• Elastic processing capacity to dynamically respond to changing throughput needs in responding to anomalies
Complete migration of all databases to PostgreSQL	• COTS licensing savings • Reduces DBA support needs and security patching overhead

Modernize DPN Storage Layer	• Product storage moved from GPFS to cloud-native blob storage (AWS S3)
	  • Significant cost savings
	  • Initial prototyping shows satisfactory performance with minimal code modifications
	• Common storage migrates to cloud-native shared file system (AWS Elastic File Service EFS)
	  • Provides HA without overhead required to manage large replicated storage cluster

Utilize Clustered Messaging Service	• Develop HA messaging system or utilize "Messaging-as-a-Service" from AWS (Amazon MQ)
Utilize Cloud-Native Monitoring and Alerting	• Initial-Implementation using legacy design of monitoring agents deployed on DPN VMs delivering messages to operations.

Modernization – Phase 3

- Potential capabilities enabled by the Cloud implementation
  - The modernization phase could leverage DPN proven data production platform
    - Provide an expanded number of enterprise data products
    - Decreases algorithm process overhead accelerating R2O cycle
  - Data delivery to an expanded user base while minimizing data egress costs
    - Prioritize Real-time products critical to NWP delivered with DPN proven low-latency and stability
    - Products are packaged and delivered as needed with all products available in S3

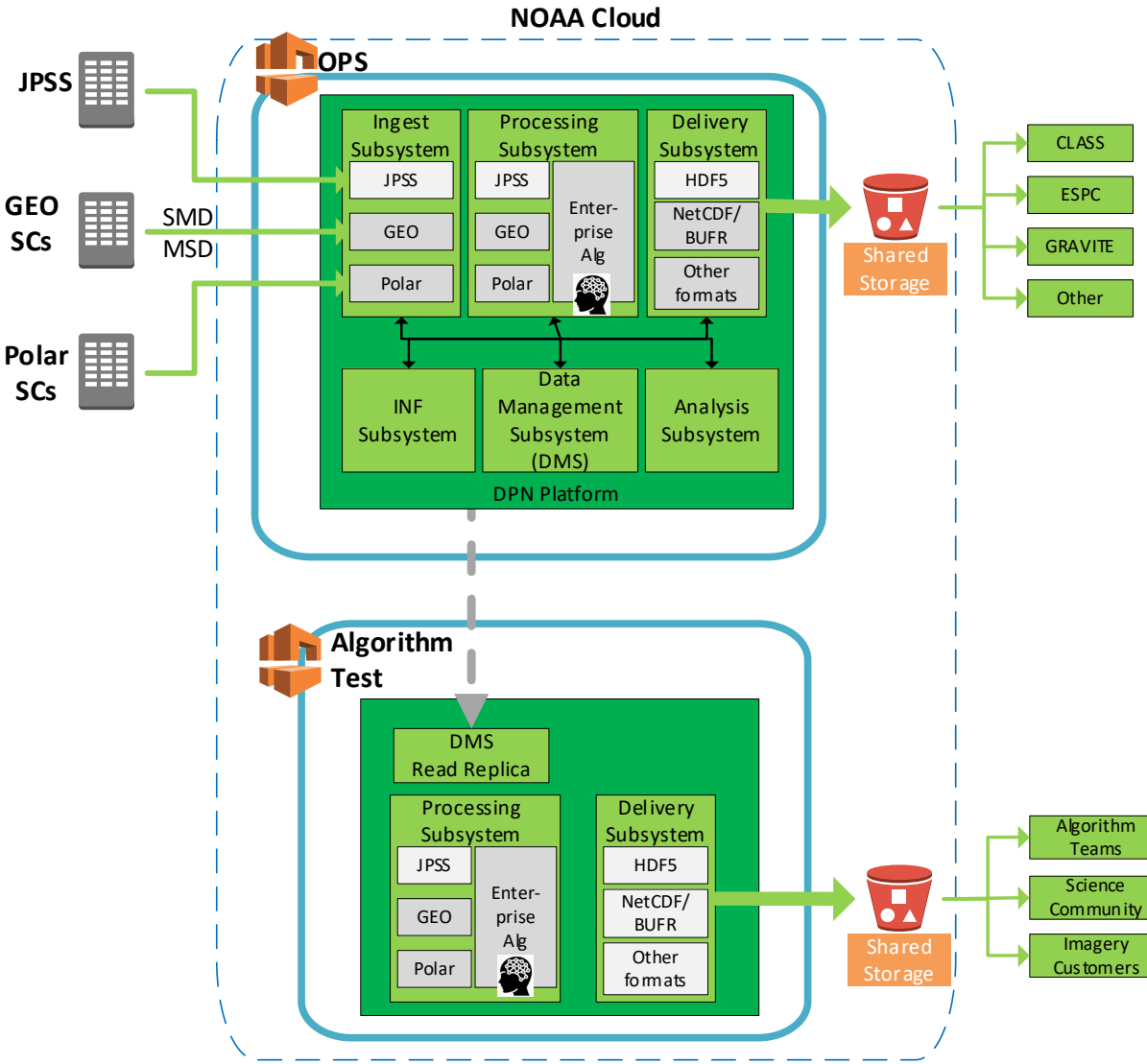
Optimization	Description

Modernize Processing Subsystem using Containerized Algorithms	• Science teams directly develop algorithms using containerized ADL and include dependencies in versioned containers
	• Run multiple algorithm versions in parallel, dependencies reside in container
	• Enterprise data product generation
	• Real-time Processing: Operational algorithms generating products
	• Off-line Processing: "Algorithm Sandbox" Evaluate updates to algorithms
	  • Executed during "back-orbits", spot-instances or serverless
	  • Eliminates need for full DP for dedicated I&T and provides faster R2O cycles

Modernize Data Delivery via Cloud-based Content Delivery Network	• Data products delivered to single cloud location (S3)
	  • Eliminate delivery of products through C3S facility to Mission Partners
	• Real-Time Delivery: Products delivered to S3 location
	  • NWP products delivered in directly ingestible format (HDF, BUFR, NetCDF, etc)
	  • Consumers who need real-time products will receive notification of new products and API to pull the data directly down to their system (S3 => SNS => SQS pipeline)
	• Off-Line Delivery:
	  • Authorized Non-Real-Time consumers will be able to request aggregation and/or packaging of products which will create a new product in S3 and notification delivered to consumer

"Lights Out" DPN decreases reliance on dedicated operations staff	• DPN is highly stable system requiring almost no human interaction to function
	  • Decreases reliance on 24x7 dedicated operators
	• Remove Java based GUIs and replace with simplified web GUI with APIs to drive DPN functions
	  • Significantly improves security posture

DPN Updates for Enterprise Algorithm Processing



- Production of new GEO/Polar mission products in parallel with JPSS OPS
  - Delivery to consumers in any format from data lake
- Potential to produce enterprise algorithms
  - Machine Learning (ML) added in-line to tag metadata with Wx features of interest
  - ML prototype developed to detect tropical cyclone activity in VIIRS granules
- "New" algorithms can be executed in parallel to operations with no impact to OPS
  - Data Driven process leveraging database/storage read-replica

CI/CD Pipeline and DevSecOps Transformation

- In parallel with DPN architectural modernization, JPSS CGS is evolving toward increased pipeline automation and DevSecOps processes
  - Expand automated testing and regression checkout
  - Automate manual parts of the deployment pipeline
  - Scrum-style development processes
- Release cycle pre-2021 was ~13 weeks – expected to move to 8 weeks in 2021 with further improvements in 2022

Summary

- JPSS CGS DPN team has outlined a path to modernization for the legacy processing system
- Some optimizations have already been prototyped and demonstrated
- Modernization Objectives:
  - Drive cloud operation costs down
  - Provide better turn around time for science and NWP products